Developing a semantic event search engine for biomedical events - Poster

Rangel, Julio C. | Kobayashi, Norio

Schlagwörter/Keywords: semantic search, PubMed, biomedical event, search engine

Developing a semantic event search engine for biomedical events

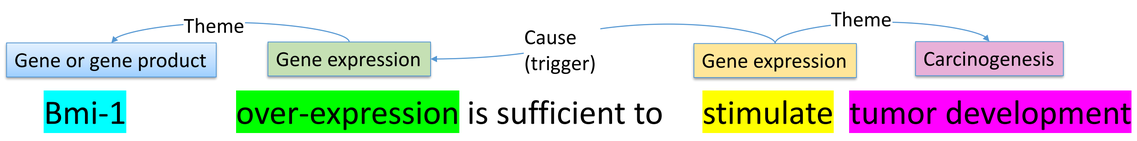
Julio C. Rangel and Norio Kobayashi

RIKEN

RIKEN Information R&D and Strategy Headquarters

INTRODUCTION

Biomedical events are modeled structures that explain biological entity connections such as proteins, medicines, and symptoms. They are normally composed of a trigger (signal word) and arguments (biological stuff), with the trigger determining the type of event and the responsibilities of its parameters. Biomedical events assist researchers in extracting useful, organized, and unambiguous information about biological entity connections from the large volume of scientific literature available.



- Biomedical events are modeled using biomedical event structures, consisting of a trigger (signal word) and arguments (biological things)
- The trigger determines the event type and specifies the roles of its parameters
- Triggers are usually nouns or verbs, arguments are proper nouns
- The role theme defines the primary subject of attention in an event, the role cause is the event's facilitator/driver.
- Scientific literature contains a large amount of relational knowledge about connections between biological entities (e.g. proteins, medications, symptoms)
- Researchers need to extract organized and clear information about biomedical events from the increasing volume of publications
- Growing number of tools to extract biomedical events, but existing search engines not utilizing this data [Li 2021].

How can we locate pertinent biomedical events in the large amount of text?

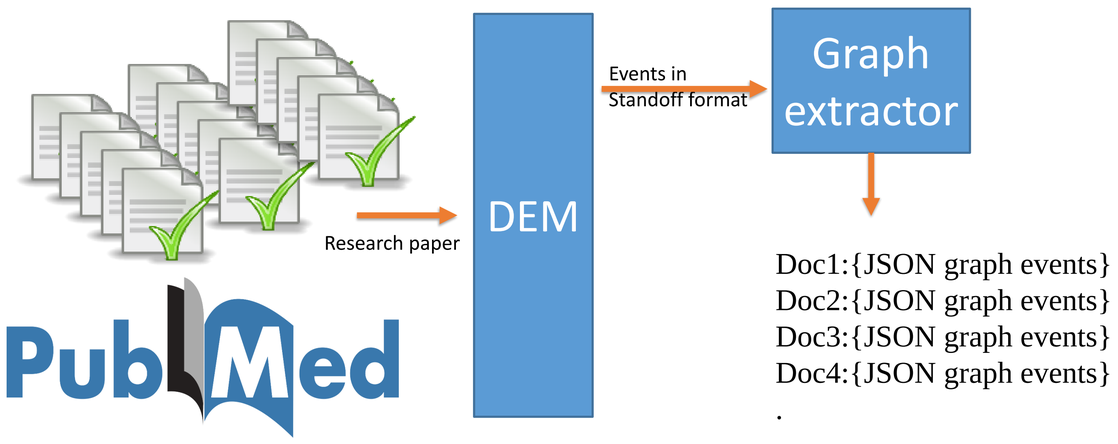
We use DeepEventMine to extract events from the PUDMED baseline and generate a Faiss index for **semantic search**. We are the first to attempt a PUBMED-wide semantic searcher for biomedical events.

OBJECTIVES

- Create a semantic search engine for biomedical events.
- Which can be accessed through a web interface
- Key feature: achieves semantic textual similarity through event embeddings
- Improves upon keyword-based search tools which suffer from keyword mismatch

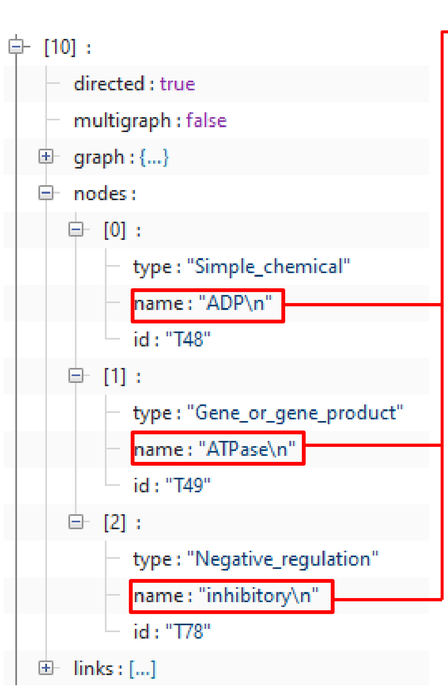
EXTRACTING BIOMEDICAL EVENTS

First, we extracted cancer genetics (CG) and infectious disease (ID) events from 34 million PubMed abstracts and titles using DeepEventMine (DEM). The standoff format is then omitted, and events are converted into a NetworkX graph and saved as smaller JSON files. In addition, we increased GPU data processing to increase DEM's speed.

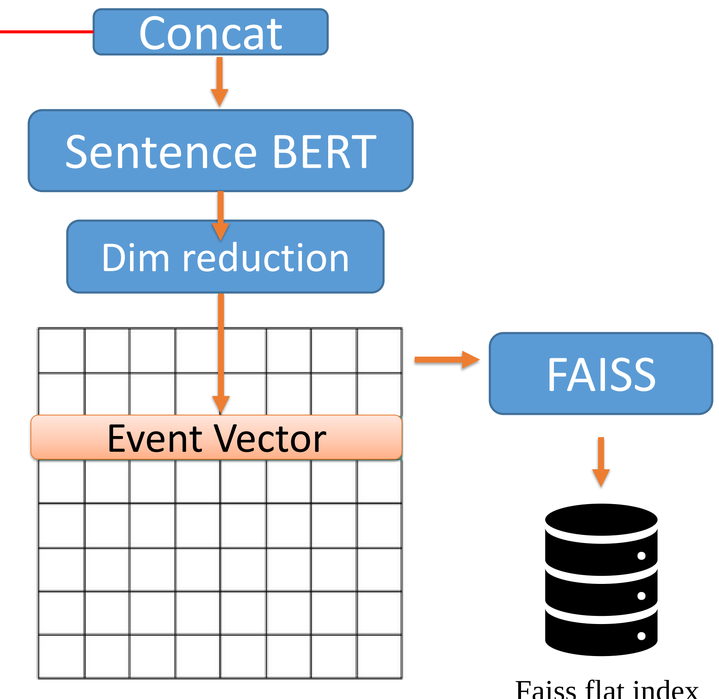


INDEX CREATION

We generated a sentence-BERT embedding for each event graph by concatenating the names of each node. In addition, index storage is reduced by decreasing the embedding dimension with BERT-whitening [Su2021a]. The resulting embeddings are converted into a Faiss flat index.



JSON event graph

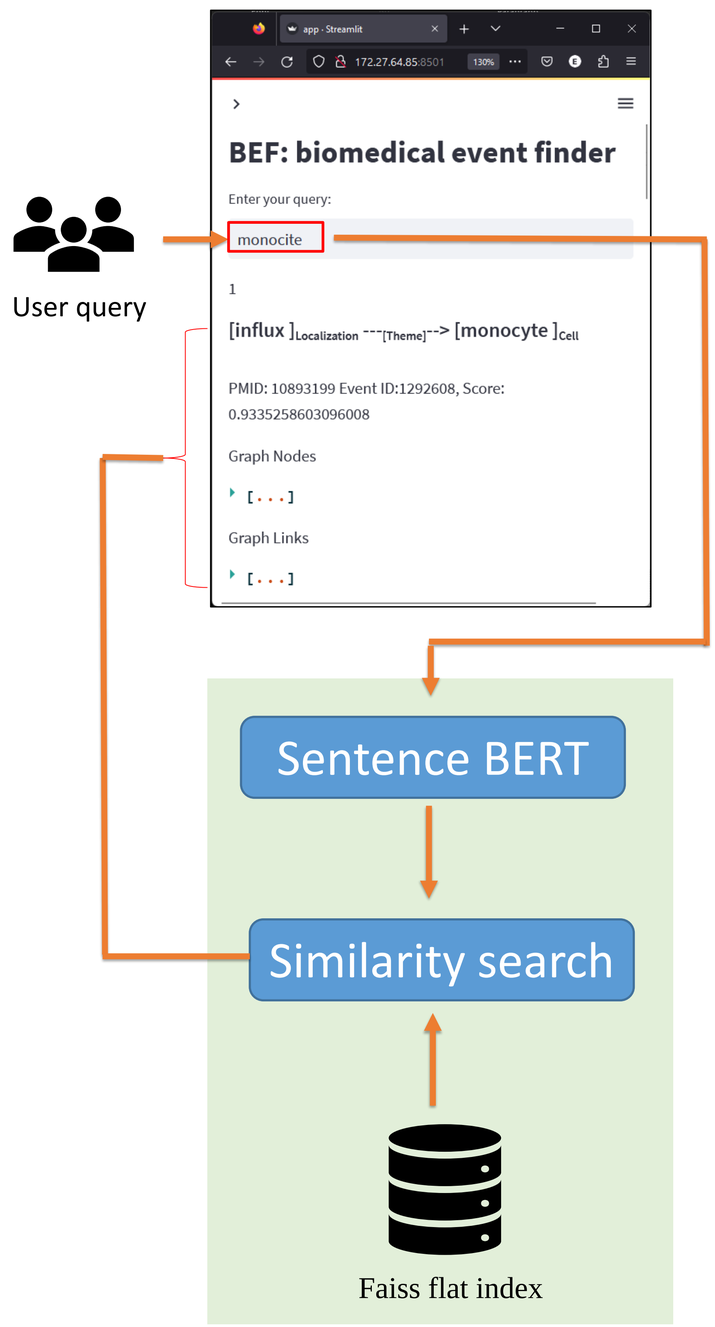


Matrix of embeddings

Faiss flat index

BIOMEDICAL EVENT FINDER (BEF): SEMANTIC SEARCH

Sentence-BERT is used by BEF to transform the query into a dense vector, after which the most similar vectors in the Faiss database are fetched.



ONGOING WORK

- Ongoing development to improve engine performance and add new functionalities
- Adding GENIA, epigenetics (EPI), pathway curation (PC), and MLEE to the list of biomedical events
- Developing and evaluating event embedding methods for improved search precision, such as GNN and DDEGK.

REFERENCES

- Li, Qian, et al. "A Compact Survey on Event Extraction: Approaches and Applications." *arXiv preprint arXiv:2107.02126* (2021).
- Su, Jianlin, et al. "Whitening sentence representations for better semantics and faster retrieval." *arXiv preprint arXiv:2103.15316* (2021).
- Zhu, Lvxing, and Haoran Zheng. "Biomedical event extraction with a novel combination strategy based on hybrid deep neural networks." *BMC bioinformatics* 21 (2020): 1-12.